Factors influencing the accuracy of age-estimates of unfamiliar faces

Article (Published Version)

Perception, 1995, volume 24, pages 1059-1073

Factors influencing the accuracy of age estimates of unfamiliar faces

Patricia A George, Graham J Hole
School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9QH, UK
Received 27 October 1994, in revised form 20 May 1995

Abstract. Factors affecting the accuracy with which adults could assess the age of unfamiliar male faces aged between 5 and 70 years were examined. In the first experiment twenty-five 'young' adult subjects, aged 16-25, and twenty-five 'old' adults, aged 51-60, were used. Each subject saw five versions of three different faces: these consisted of an original version of each face and four manipulated versions of it. The manipulations consisted of mirror reversal, pseudo-cardioidal strain, thresholding, and elimination of all but the internal features of the face. The second experiment was similar except that a between-subjects design was used: each subject saw three faces for each age category of target face, but was exposed to only a single type of manipulation (plus a set of 'original' faces which were identical for all groups, so that the comparability of the different groups in age estimation could be checked).

Results from both experiments were similar. Age estimates for unmanipulated 'original' faces were highly accurate, although subjects were most accurate with target faces that were within their own age range. Results for the manipulated faces implied that the importance of cardioidal strain as a necessary and sufficient cue to age may have been overestimated in previous reports: subjects' age estimates were accurate when cardioidal strain was absent from the stimulus, and poor when cardioidal strain was the only cue available.

1 Introduction

Bruce and Young's (1986) model of face perception highlights the fact that faces provide a variety of different types of information: a face tells us about a person's identity, gender, and emotional state, as well as aiding speech comprehension. Faces provide information concerning yet another attribute of their owners, but one about which Bruce and Young's model has relatively little to say: a person's age. Given that we can assess the age of a face, what information is being extracted in order to enable us to do so, and what is its relationship to the information used to establish, for example, emotional expression and individual identity? This is an interesting issue both from practical and from theoretical points of view. Since one of the characteristics most readily used to describe someone is their apparent age (Shepherd et al 1981), there is an obvious practical application of research in this area: for example, it is useful to know how much confidence can be placed in eyewitnesses' estimates of suspects' ages.

Ageing affects the appearance of a face in diverse ways. From childhood to adulthood there are changes in head shape, with the face becoming markedly larger in proportion to the cranium (Enlow 1982). In adulthood wrinkling and creasing occur in response to stresses and skin-texture changes. With advancing years hair may change colour, become thin or recede, and teeth may disappear. Each of these characteristics may provide visual information about a person's age (Mark et al 1980), but only one cue has been the subject of detailed experimental study. 'Cardioidal strain' is a mathematical transformation that models the physical changes that occur as the infant's face and cranium grow until adulthood. A baby's face is small and overshadowed by a large forehead. Through growth and ageing the face comes to occupy a larger part of the head, and the internal features (eyes, nose, and mouth) effectively move upwards on the head. During the past twenty years numerous studies have demonstrated that manipulations of the level of cardioidal strain applied to a face affect subjects' judgments of the age of the face (eg Pittenger and Shaw 1975; Pittenger et al 1979; Mark and Todd 1983, 1985; Bruce et al 1989). Faces appear older if the strain transformation is applied in one direction, and younger if it is applied in the opposite direction. ['Pseudo-strain', in terms of merely moving the internal features within the face outline, has similar effects, as shown by Bruce et al (1991).] Working within a neo-Gibsonian framework, Pittenger and his colleagues have suggested that cardioidal strain level (or something similar) is a perceptual 'invariant' which constitutes the primary perceptual information for age estimation.

Previous research has shown that cardioidal strain may be a sufficient cue for age perception, in that subjects can make rank-order judgments of age solely on the basis of the profile of the head and face, with all other cues to age removed. However, cardioidal strain is equally certainly not the sole cue to age. Since we are able to discriminate between different-aged adults, this implies that cues other than cardioidal strain must be used for perception of adults' ages.

Other cues to age have been studied relatively little. Mark et al (1980) investigated the extent to which adult subjects could use changes in head shape and the amount of facial wrinkling as guides to age. Subjects used both cues, although wrinkling was found to have a more pronounced effect than head shape. The most accurate age estimates were obtained when the two manipulations were congruent with each other; the effect on age perception of changes either in wrinkling or in head shape depended on the state of the other characteristic. Montepare and McArthur (1986) found that preschool children were able to use a number of cues to identify the age category to which a face belonged: although two of these (profile shape and the vertical location of features within a frontally viewed face) can be simulated by cardioidal strain, the most salient cue to age, for children, seemed to be the degree of facial wrinkling—a finding which supported Jones and Smith's (1984) earlier observations. Such studies suggest that in order to obtain a fuller understanding of how people estimate age it is necessary to consider cues to age other than cardioidal strain.

A second problem with this area of research lies in the artificiality of the stimuli used. In most of the studies manipulating cardioidal strain line drawings of face profiles have been used. Mark et al (1980) used tracings from an artist's line drawings of heads in profile, which were intended to represent different ages. The most realistic stimuli used so far have been Bruce et al's (1991) 'Mac-a-Mug' composites. All of these stimuli are relatively impoverished in terms of the cues that they provide as a basis for age estimation, and it is conceivable that such impoverishment might encourage subjects to use whatever cues remain (such as cardioidal strain) as a basis for their judgments to a greater extent than they might if a wider range of cues were available.

Finally, with the exception of developmental studies of age perception, few researchers have considered the age of the subject as a variable affecting age estimates. To our knowledge, in all previous studies of age perception young adults (undergraduates) have been used as subjects. We wanted to extend the generality of our results to older, non-student, populations.

In response to these gaps in the present literature on face perception and age perception, the following two experiments were devised. Subjects were asked to estimate the age (in years) of computer-manipulated faces. These were prints obtained from full-face black-and-white photographs, rather than schematic faces or line-drawn profiles. The faces varied in age (from 5 to 70 years), and each face was presented in four different versions which either possessed or lacked particular possible cues to age. One version was an unaltered image. In a second version the internal features were displaced downwards within the face outline to produce pseudo-cardioidal strain (in a direction which should make the face appear 'younger'). A third version showed only the internal features so that cardioidal strain was unavailable to the subjects, at least in any direct manner. In a fourth version, the face was complete but was thresholded so that the features and hair appeared as black blobs on a white background—a manipulation which preserved information about cardioidal strain, but removed any detailed information about hair, skin texture, and wrinkling. As a way of checking on the consistency with which subjects made their age judgments, a final version consisted of a mirror reversal of the original face.

Since the subjects' own ages may play a part in their estimates of other peoples' ages, we looked at the effects of these manipulations on the accuracy of age estimates made by young adults (aged 16 – 25 years) and older adults (aged 51 – 60 years).

Our predictions were as follows: if cardioidal strain is the primary cue for accurate age estimation then the pseudo-cardioidal strain version should appear younger than the original unaltered face (thus replicating, with more realistic stimuli, previous researchers' findings with schematic faces, line-drawn profiles, and 'Mac-a-Mug' faces). Subjects should be as accurate in their judgments of the age of the thresholded faces as they are in their judgments of the original faces, since the former preserve the strain information present in the latter. Conversely, subjects should perform less accurately with faces consisting only of internal features since these provide little information about the level of cardioidal strain. There is no reason to expect age estimates of the mirror-reversed faces to differ from those of the original versions, since all cues (strain level, skin texture, etc) are identical. If experience with faces is involved in age estimation, one might expect young and old subjects to be more accurate in estimating the age of faces which were similar in age to their own.

2 Experiment 1

2.1 *Method*

2.1.1 *Design*. There were three independent variables. First, the age of the target face—there were seven age categories of stimulus photograph: 5 – 10, 15 – 20, 25 – 30, 35 – 40, 45 – 50, 55 – 60, and 65 – 70 years. For each age category, photographs of three different individuals were used, so that a total of twenty-one different target faces were used.

Second, the version of the target face—five different versions of each target face were presented: 'Original', 'Reversal', 'Moved', 'Thresholded', and 'Features' (see below for definitions).

Third, the age of the subject: half the subjects were aged between 16 and 25, and half were aged between 51 and 60 years.

All subjects were shown all permutations of age and version of the target faces (giving rise to 105 images—seven ages × five versions × three target individuals per age). The dependent variable was the subject's estimate of the age of each target face, in years.

2.1.2 *Subjects*. There were forty-eight subjects, twenty-four in each age group. Half of the subjects in each age group were male. All subjects were unaware of the purposes of the experiment.

2.1.3 *Apparatus and materials*. A total of 21 photographs of males aged between 5 and 70 years were taken (3 photographs for each of the seven age categories of target face; the adults were clean-shaven). These were scanned into Ansel Image Editing Software on an IBM PC by the use of a Logitech 256 grey-scale scanner set for 256 grey levels and a resolution of 200 dots per inch. For each of the 21 original photographs a set of five versions was produced. Each set consisted of the original face (henceforth referred to as 'Original'), plus four variants: mirror reversal ('Reversed'), feature displacement ('Moved'), thresholded ('Threshold'), and features only ('Features').

After manipulation, the resulting images were output as 105 A4-sized monochrome prints, via an Epson FX-80 9 pin printer. The different versions were as follows:

(a) The Original version was an unaltered print from the original scanned photograph.

(b) The Reversed version was the original image mirror-reversed. This was a control condition introduced to check on the consistency of subjects' age estimates.

(c) The Moved image was the result of a manipulation made to the original face to mimic cardioidal strain level in a direction which should make the face appear younger. The features of the face (eyebrows, eyes, nose, and mouth) were moved downwards by approximately one third of an inch. The resulting gaps were filled in with the appropriate grey levels to make sure that the moved feature blended in with the rest of the face.

(d) The Threshold image was a manipulation of the original face which was intended to remove all detailed skin-texture cues, while retaining information about cardioidal strain level. The grey levels in the original faces were thresholded, by taking the midpoint of the grey-level range and setting all grey levels above this value to white, and all values below it to black. This produced images in which features appeared as black blobs on an otherwise-white face. (The result is effectively a high-contrast, low-spatial-frequency image, although no formal measurements of the spatial-frequency content of the face were made).

(e) The Features image was one in which we intended to remove such cues to cardioidal strain as face shape and hair. All of the face beyond a rectangular area containing the internal face features (eyebrows, eyes, nose, and mouth) was deleted, to leave the internal features surrounded by white background.

2.1.4 *Procedure.* Each subjects was shown all 105 images sequentially. A different random order was used for each subject, subject to the constraint that no target face came within 7 images of any of the other versions of the same face.

Subjects were told that the experimenter was interested in finding out about how people decide how old a face is. For each face, subjects were asked to write down the apparent age in years. Subjects were told that, although some faces might seem to recur in the series, none of the faces would be identical to any other shown; they were to rate each face independently of any of the others. They were asked to take as long as they considered necessary to make that judgment.

2.2 *Results*

2.2.1 *Treatment of results.* The raw data consisted of each subject's estimated ages for the 105 faces: for each permutation of seven age categories and five versions of target face, there were three different faces to be judged by each subject. Each subject's age estimates for these three faces were averaged together, so that every subject provided 35 mean age estimates (one for each permutation of age category and version).

To address the question of how accurate subjects' age estimates were, individual subjects means were averaged together to give a mean estimated age for each of the seven age categories. This was done separately for each of the five versions of face. Figures 1a and 1b show, separately for the two age groups of subject, these mean age estimates plotted against the actual mean chronological age of the target faces within each age category. These graphs emphasise the overall accuracy of subjects' age estimates: perfectly accurate age estimates would fall on a line at 45° to the axis; overestimates fall above and underestimates below this line, respectively. Figures 2a and 2b show the same data in a different format, in order to emphasise the differences between the experimental conditions: the age estimates for the Reversed, Moved, Features, and Threshold versions are expressed as percentages of the age estimates for the Original versions.

For each of the seven age categories of face, a two-way analysis of variance was performed (two levels of subject age × five levels of target-face version: the first variable comprised independent measures and the second variable was within subjects). The results of these ANOVAs are summarised in table 1. All ANOVAs showed highly significant main effects of face version. There were highly significant main effects of subject age for all age categories of target face except for the 25–30-year-old age category, where the effect of subject age was only marginally significant ($F_{1,46} = 3.01$, $p < 0.10$). The interaction between subject age and target-face version was significant for all age categories of target face.

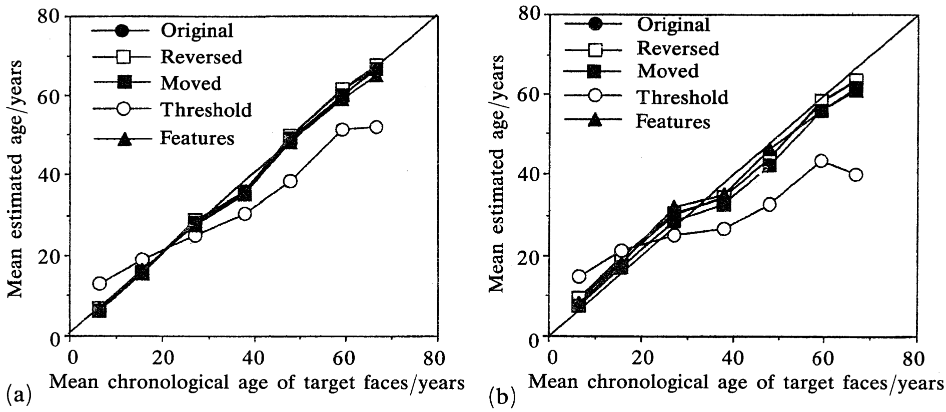

Figure 1. Data from experiment 1: mean age estimates for the seven age categories of target face, as a function of face version, plotted against the mean chronological (ie true) ages of the target faces representing each age category. (a) Estimates made by subjects aged 16 to 25 years; (b) the estimates by subjects aged 51 to 60 years.

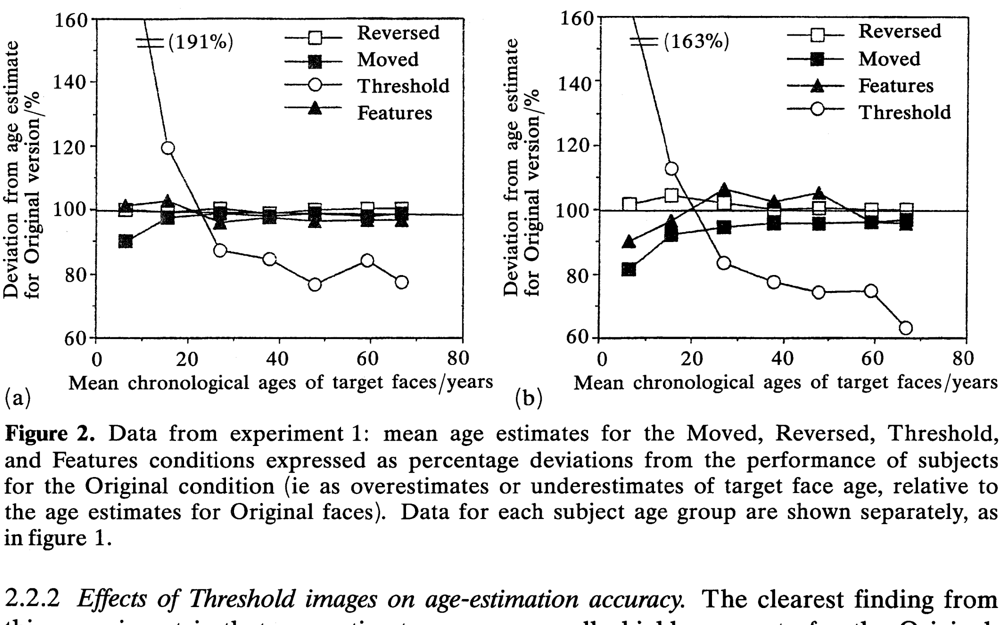

Figure 2. Data from experiment 1: mean age estimates for the Moved, Reversed, Threshold, and Features conditions expressed as percentage deviations from the performance of subjects for the Original condition (ie as overestimates or underestimates of target face age, relative to the age estimates for Original faces). Data for each subject age group are shown separately, as in figure 1.

2.2.2 *Effects of Threshold images on age-estimation accuracy.* The clearest finding from this experiment is that age estimates were generally highly accurate for the Original, Reversed, Moved, and Features versions, but performance was markedly poorer for the Threshold faces. This is shown clearly in figures 1a and 1b, and this interpretation was confirmed by a posteriori Newman–Keuls tests. Both age groups of subjects

subjects considered the 5 - 10-year-old and 15 - 20-year-old Threshold faces to be significantly older than the other versions ($p < 0.01$ for all comparisons between Threshold faces and other versions, at both ages), whereas Threshold faces in the remaining five age categories were considered to be younger than the other versions ($p < 0.01$ for all comparisons between Threshold faces and other versions, at all five ages). Compared with the actual ages of the target faces, the younger subjects overestimated the age of the Threshold version in the 5 - 10-year-old and 15 - 20-year-old age categories by up to 6 years, and underestimated the ages of the older faces by as much as 15 years. Older subjects overestimated the age of the Threshold faces in the 5 - 10-year-old and 15 - 20-year-old age categories by up to 9 years, and underestimated the ages of the older faces by as much as 22 years.

In the case of the Original, Reversed, Moved, and Features versions, the older subjects were poorer than the younger subjects at judging age: whereas the younger subjects' estimates were within 2 years of the actual ages of the target faces, the older subjects' estimates were generally within no more than about 5 years, even with the Original faces. For all versions, the older subjects' performance was least accurate with the youngest and oldest age categories of target face, older subjects tending to overestimate the age of young faces and underestimate the age of faces similar in age to their own!

Table 1. Summary of results of ANOVAs conducted on data from experiment 1, for each age category (in years) of target face. F ratios for main effects of face version and the interaction face version × subject age group are for 4,184 degrees of freedom, and for main effects of subject age group are for 1,46 degrees of freedom.

Age category of target face	F ratio main effect		F ratio interaction face version × subject age group
	face version	subject age group	
5 - 10	262.41, $p < 0.001$	21.81, $p < 0.001$	2.37, $p = 0.05$
15 - 20	26.73, $p < 0.001$	30.41, $p < 0.001$	2.66, $p < 0.05$
25 - 30	59.50, $p < 0.001$	3.01, ns	11.79, $p < 0.001$
35 - 40	114.67, $p < 0.001$	4.77, $p < 0.05$	6.99, $p < 0.001$
45 - 50	182.34, $p < 0.001$	31.13, $p < 0.001$	6.75, $p < 0.001$
55 - 60	201.49, $p < 0.001$	16.19, $p < 0.001$	8.88, $p < 0.001$
65 - 70	248.25, $p < 0.001$	34.10, $p < 0.001$	11.19, $p < 0.001$

2.2.3 *Effects of pseudo-cardioidal strain on age-estimation accuracy.* For younger subjects, estimates of the ages of the Moved pictures were little different from the true ages of the target faces concerned: thus there was little effect of pseudo-cardioidal strain on absolute age judgments. For each age category, the younger subjects' age estimates for the Original and Moved target faces were compared by the use of t-tests. (Whereas the specific comparisons in the previous and subsequent sections were a posteriori and therefore warranted use of an a posteriori test, we had planned to make the comparisons in this section before the experiment was conducted.)

Although the Moved faces were rated as younger than the Original faces in all age categories, the only statistically significant comparison was for the 5 - 10-year-old faces ($t_{23} = 6.88$, p < 0.0001, two-tailed test). The younger subjects consistently judged the Moved faces to be a year or less younger than the Original faces: thus pseudo-cardioidal strain appears to have had progressively more effect on age judgments of younger target faces, since a year is a proportionately large difference at the low end of our age-category range (5 - 10 years) and a negligible difference at the high end (65 - 70 years).

For older subjects, the effects of pseudo-cardioidal strain in the Moved pictures were more pronounced, though still relatively small in absolute terms. For each age category, a t test was used to compare the older subjects' age estimates for the Original and Moved versions. Subjects consistently judged the Original face as being approximately 2 years older than their Moved counterparts. As with the younger subjects, this implies that pseudo-cardioidal strain was having a proportionately greater effect at the younger ages. All Moved versus Original comparisons were statistically significant (smallest t was for Moved versus Original in the 25–30-year-old age category: $t_{23} = 2.61$, $p < 0.02$, two-tailed test).

2.2.4 *Effects of Features condition on age-estimation accuracy.* Newman–Keuls tests revealed relatively few differences between subjects' performance on the Features faces and their performance on the Original versions, and those that were found were rather inconsistent. For younger subjects, the only statistically significant differences were that the Features versions were rated as significantly younger than their Original counterparts in the 25–30-year-old and 55–60-year-old age categories (Newman–Keuls test, $p < 0.05$ in both cases). For older subjects, the Features manipulation seemed to have more effect on age estimates: these versions were rated as significantly younger than the Original versions in the 5–10, 55–60, and 65–70 years age categories ($p < 0.01$ for the 55–60-year-old age category and $p < 0.05$ for the other two comparisons). However, for the 25–30-year-old and 45–50-year-old age categories, the Features versions were rated as significantly older than their Original counterparts ($p < 0.01$ in both cases), and for the 15–20-year-old and 35–40-year-old age categories the differences did not approach significance.

2.2.5 *Accuracy of age estimates in relation to subjects' own age.* To investigate the hypothesis that subjects would be more accurate at estimating the age of faces similar in age to themselves, direct comparisons were made between the performance of the younger and older subjects. A two-way mixed ANOVA (two levels of subject age group × seven levels of target-face age category) was used to compare the younger and older subjects' performance on the Original faces, for each age category of target face. This revealed significant main effects of subject age group ($F_{1,46} = 6.87$, $p < 0.01$) and target-face age category ($F_{6,276} = 2320.30$; $p < 0.0001$) plus a significant interaction between subject age group and target face age category ($F_{6,276} = 14.32$, $p < 0.001$).

These results stem from the fact that for the two youngest age categories, the younger subjects' age estimates were more accurate than were the older subjects' estimates. For the three oldest age categories, the reverse was true: the older subjects' estimates were more accurate than were the younger subjects' estimates.

2.3 *Summary and conclusions*

Although subjects were good at judging the ages of unfamiliar faces in their Original and Reversed versions, the Threshold manipulation significantly reduced the accuracy of their estimates. In contrast, the Features manipulation did not significantly impair performance. One would expect the opposite results if cardioidal strain were the principal cue for age perception, since the Threshold faces retain the cardioidal strain information of the original faces and the Features faces do not.

Given the importance assigned to the cardioidal strain manipulation in previous research, we were surprised to find little effect of the Moved manipulation on the younger subjects' age estimates. However, the older subjects were somewhat more influenced by the Moved manipulation: application of pseudo-cardioidal strain to target faces caused them to appear to be younger than they actually were.

Subjects were better at judging the age of faces which were similar in age to their own. Speculation on the possible reasons for these effects will be deferred until section 4.

3 Experiment 2

3.1 *Method*

A potential methodological problem with experiment 1 stems from the repeated-measures aspects of its design: by showing each subject all versions of a particular face, we inadvertently gave them the opportunity to sometimes use the Original version as a basis for their estimates of the ages of the manipulated versions. For example, a subject who was unable to directly estimate the age of a Features version with any accuracy might nevertheless provide an accurate estimate of its age if they were able to recall their estimate of the age of the Original version of that face which they had seen previously.

There are two reasons why this is unlikely to be the whole explanation of our obtained results. First, subjects saw the different versions of a face in different random orders: in many cases one or more manipulated versions were encountered before the Original version of a face was seen. In these cases it would obviously be impossible for the Original version to serve as a guide for age estimates for the manipulated versions. Second, if cross-cueing of this kind was occurring, it is hard to explain why subjects were so poor at estimating the ages of the Threshold faces, and why their age estimates of the Moved faces were influenced by the pseudo-cardioidal strain manipulation.

However, to control for the problem of cross-cueing between different target-face versions, the previous study was repeated with a different design. The accuracy of age estimates for the Moved, Threshold, Features, and Original versions was compared, with a different group of subjects for each of these conditions. The same fourteen target faces (two for each of seven age categories) were presented to all subjects, but a given subject saw only one type of manipulation applied to all fourteen faces. For example, subjects participating in the Moved condition saw all faces after they had been subjected to pseudo-cardioidal strain, whereas subjects participating in the Threshold condition saw the same faces, but in their thresholded versions. To ensure that the four groups of subjects were comparable in age-estimation ability for normal faces, all subjects also saw a third face for each age category, in its Original version only. (To distinguish this third set of faces from the other Original faces, they will henceforth be referred to as the 'Control Original' faces.)

In line with the previous experiment, we would predict that the four groups would differ significantly in their age estimates: the Moved group subjects should consistently underestimate the ages of the faces that they see, the Threshold group subjects should make less accurate estimates than the other groups, and the Features group subjects should produce a performance similar to that of the Original group. The latter should produce the most accurate age estimates. All four groups should not differ in their estimates of the ages of the Control Original faces that they see in common.

3.1.1 *Subjects.* Eighty subjects participated in this experiment. Forty were 'young' (aged 16 – 25 years) and forty were 'old' (aged 51 – 60 years). Half of the subjects in each group were male.

3.1.2 *Apparatus.* 84 stimulus faces were prepared, as for experiment 1. These consisted of four versions (Original, Moved, Threshold, and Features) of three different faces for each of seven different age categories (the same age categories and versions of face as were used in experiment 1).

3.1.3 *Procedure.* There were four groups of twenty subjects, representing each permutation of age (old or young) and target-face version (Moved, Threshold, Features, or Original). Each subject saw 21 images, in a different random order, and estimated the age of each face in turn.

For each age category, there were three different target faces: A, B, and C. Depending on whether a subject was in the Original, Moved, Threshold, or Features condition, he or she saw faces A and B in this version only. Additionally, all subjects saw face C, which was always presented in its Control Original version.

Each subject thus saw 14 target images (two different faces for each of the seven age categories) in an Original, Moved, Threshold, or Features version, plus 7 faces (one for each age category) which were in the Control Original version. This procedure enabled us to present each face in all four versions and obtain estimates of its age which were uncontaminated by subjects having knowledge of what the other versions of a face looked like, because any particular individual saw only two types of image (Control Original and one other) and never encountered the same face twice.

3.2 *Results*

The data from an individual subject consisted of three age estimates (in years) of 3 different images for each of seven age categories. Two of these faces had received the same manipulation (Original, Moved, Threshold, or Features), and the third was a Control Original face. (Statistical analysis of the data obtained from the Control Original faces, seen by all subjects, confirmed that the four subject groups were indeed comparable in age-estimation performance to begin with. Consequently, data from these faces will not be discussed further.)

The age estimates for the two faces which had received the same manipulation were averaged together. Subjects in the Original-images group saw three Original faces, two 'experimental' and one 'control'; each average was between those pairs of faces in each age category which were presented in manipulated versions to the other subject groups. This provided 7 mean age estimates (one for each age category) for each subject: these were used as the raw data in the following analyses.

As with experiment 1, these data are shown plotted in two different ways, to emphasise different aspects. Figures 3a and 3b show the mean age estimates for the four different versions of face, plotted against the corresponding chronological ages of the target faces, and taking into account the subjects' own age. Figures 4a and 4b express the age estimates for the Moved, Threshold, and Features conditions in terms of their deviations from the performance of subjects in the Original-images condition.

It is apparent from these graphs that, as in experiment 1, subjects were generally very accurate at estimating the ages of the Original, Moved, and Features faces,

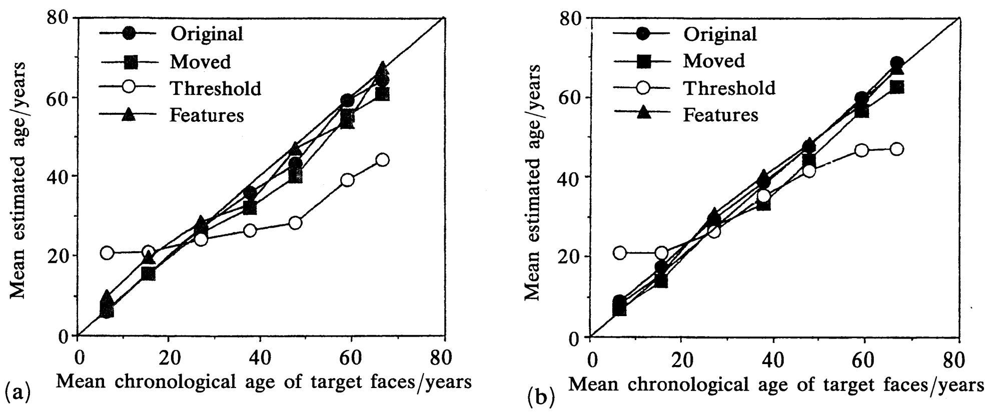

Figure 3. Data from experiment 2: mean age estimates for the seven age categories of target face, as a function of face version, plotted against the mean chronological ages of the target faces. (a) Estimates made by subjects aged 16 to 25 years; (b) estimates by subjects aged 51 to 60 years.

and considerably poorer at judging the age of the Threshold versions. This impression was confirmed by statistical analysis. For each age category of face, a two-way independent-measures analysis of variance was performed (two levels of subject age group by four levels of face version). Table 2 summarises the principal results from these analyses. All seven ANOVAs showed a highly significant main effect of face version. The main effect of subject age group was highly significant in four of the ANOVAs (for the 25–30, 35–40, 45–50, and 55–60 year age categories). For the 5–10, 15–20, and 65–70 year age categories of target face, the effect of subject age group failed to reach statistical significance, although there was a definite trend towards significance. The interaction between subject age group and face version was statistically significant for the 15–20, 35–40, and 45–50 year age categories of target face and not significant for the rest.

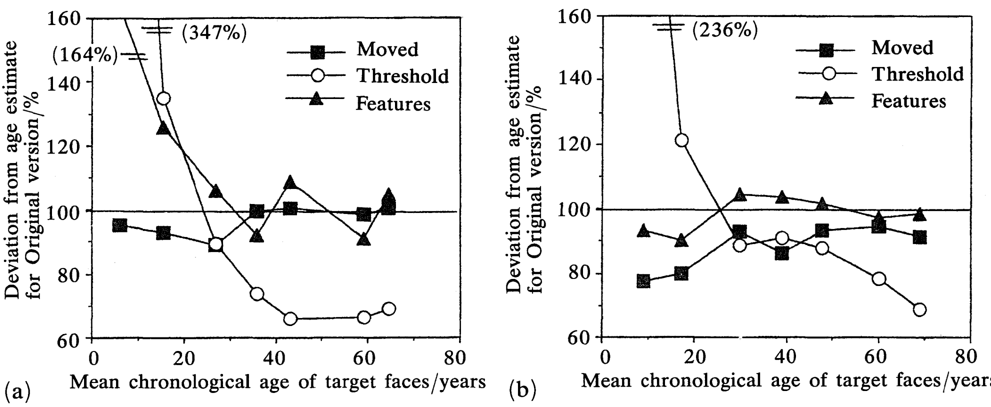

Figure 4. Data from experiment 2: mean age estimates for (a) subjects aged 16 to 25 years and (b) subjects aged 51 to 60 years in the Moved, Threshold, and Features conditions as percentage deviations from the performance of subjects for the Original condition.

Table 2. Summary of results of ANOVAs conducted on data from experiment 2, for each age category (in years) of target face. *F* ratios for main effects of face version and the interaction face version × subject age group are for 3,72 degrees of freedom, and for main effects of subject age group are for 1,72 degrees of freedom.

Age category of target face	*F* ratio main effect		*F* ratio interaction face version × subject age group
	face version	subject age group	
5–10	115.90, $p < 0.001$	7.20, ns	2.27, ns
15–20	24.42, $p < 0.001$	3.58, ns	5.16, $p < 0.005$
25–30	8.19, $p < 0.001$	11.09, $p < 0.001$	0.06, ns
35–40	12.16, $p < 0.001$	32.10, $p < 0.001$	3.90, $p < 0.02$
45–50	23.56, $p < 0.001$	27.24, $p < 0.001$	5.39, $p < 0.005$
55–60	30.24, $p < 0.001$	7.06, $p < 0.01$	1.46, ns
65–70	60.07, $p < 0.001$	2.97, ns	0.49, ns

3.2.1 *Effects of Threshold images on age-estimation accuracy.* As in experiment 1, Newman–Keuls a posteriori tests revealed that the significant results of face version occurred because subjects' age estimates were markedly less accurate for the Threshold faces than for the other conditions: there was little difference in performance on the other three versions (Moved, Original, and Features).

Also as in experiment 1, subjects overestimated the age of young target faces when they were thresholded, and underestimated the age of thresholded older faces.

In comparison with the Original versions, younger subjects significantly overestimated the age of the Threshold faces in the 5-10 and 15-20 year age categories ($p < 0.01$ in both cases), and significantly underestimated the age of Threshold faces in the 35-40, 45-50, 55-60, and 65-70 year age categories ($p < 0.01$ for all except the 35-40 year age category, where $p < 0.05$).

Older subjects showed a similar pattern: they judged the Threshold versions to be significantly older than their Original counterparts in the 5-10 and 15-20 year age categories ($p < 0.01$ in both cases) and significantly younger than the Original versions for the remaining five age categories ($p < 0.01$ in all cases).

3.2.2 *Effects of pseudo-cardioidal strain on age-estimation accuracy.* For each age category in turn, a t test was used to compare the age estimates for Original and Moved faces. This was done separately for younger and older subjects.

For younger subjects there were no statistically significant effects of pseudo-cardioidal strain for any of the seven age categories; however, with the exception of the 5-10-year-old category (in which the Moved faces were judged as appearing slightly older than their Original versions) there was a consistent tendency to give lower age estimates to the Moved faces.

In contrast, the older subjects consistently judged the Moved faces to be younger than the Original versions. This was a trend for all age categories, and was statistically significant for five of them: for the 5-10-year-old age category $t = 2.65$, $p < 0.02$, for the 15-20-year-old age category $t = 3.27$, $p < 0.005$, for the 45-50-year-old age category $t = 5.37$, $p < 0.0001$, and for the 55-60-year-old age category $t = 2.98$, $p < 0.01$. For the 25-30-year-old age category the results were marginally significant at $t = 2.05$, $p < 0.06$. For the 65-70-year-old age category $t = 1.91$ (ns) (all tests were two tailed, with 18 degrees of freedom).

3.2.3 *Effects of Features condition on age-estimation accuracy.* As mentioned earlier, there was little effect of the Features manipulation on age-estimation accuracy—the only discrepancies found between estimates for the Features and Original versions were that younger subjects significantly rated the Features versions as older than the corresponding Original faces in the 5-10-year-old and 15-20-year-old age categories (Newman-Keuls $p < 0.01$). There were no significant differences between age estimates for the Original and Features versions made by older subjects.

3.2.4 *Accuracy of age estimates in relation to subjects' own age.* Experiment 1 provided some support for the suggestion that subjects were more accurate at estimating the age of faces similar in age to themselves. This issue was examined in the same way in the present experiment. A two-way mixed ANOVA was used to compare the younger and older subjects' age estimates for the Original faces in each of the seven age categories of target face. This revealed highly significant effects of subject age group ($F_{1,78} = 20.41$, $p < 0.001$) and target face age category ($F_{6,468} = 2154.68$, $p < 0.0001$), but no interaction between age of subject and age category of target face ($F_{6,468} = 1.24$, ns). Younger subjects were more accurate than were older subjects at estimating the age of younger faces and, conversely, older subjects were more accurate at estimating the age of older subjects. However, for all age categories of target face the younger subjects provided lower estimates of the ages of Original faces than did the older subjects. Consequently one cannot rule out the possibility that these differences in accuracy are merely related to different response biases among younger and older subjects, rather than being due to differential treatment of target faces more similar in age to that of their own age group.

4 Discussion

In both experiments, both age groups of subjects were generally highly accurate in their estimates of the ages of unfamiliar faces, especially when these faces were presented in their Original versions. The effects of the experimental manipulations employed in these two experiments provide some insight into the nature of the cues used in order to assess age. The Features versions retained the skin texture and internal features of the Original faces, but reduced the use of cardioidal strain as a cue to age by removing information about the shape of the head, the hair, and the position of the facial features within the head outline. In contrast, the Threshold faces retained information about pseudo-cardioidal strain since the spatial arrangement of the features and their position on the skull were clearly visible, but all cues to skin and hair texture were eliminated. In both experiments subjects were able to accurately estimate the age of the Features faces, but found it difficult to assess the age of the Threshold versions. These results suggest that the importance of cardioidal strain as a cue to age has been overstated in previous research. In fact, precisely the opposite results to those obtained in experiments 1 and 2 would be expected if cardioidal strain were the primary cue for age perception.

The accurate performance with the Features faces suggests that age can be estimated accurately when direct information about cardioidal strain is absent: wrinkles, bags, and sagging are presumably highly satisfactory by themselves as cues to age. An alternative explanation, which should not be discounted, is that the level of cardioidal strain can somehow be inferred solely on the basis of the disposition of the internal features of a face—that subjects are able to work out what the shape of the entire head might be merely on the basis of the features. However, even if this were so, one would be moving away from the original conception of cardioidal strain as a perceptual 'invariant' in the Gibsonian sense.

The poor performance with the Threshold faces reinforces the interpretation that facial surface information is important by showing that removal of information about skin texture significantly impairs subjects' ability to judge age, and demonstrating that cardioidal strain information in isolation is a rather imprecise cue to age. Although subjects in both experiments were able to assess the age of Threshold faces in general terms, their estimates were very crude compared with the estimates obtained with the other versions of the faces. The overall pattern of results for this version shows a 'flattening' effect on age estimates. Younger faces were estimated as being older than they actually were, and older faces were estimated as being younger. The rejuvenating effect of thresholding on older faces is presumably because it eliminates information about wrinkling by removing high-spatial-frequency information from the face (a trick long known to photographers of ageing Hollywood starlets).

An important point to note is that in previous studies authors have relied on asking subjects to rank order a set of stimuli in terms of their apparent ages. For example, Mark and Todd (1985) and Bruce et al (1989) obtained their data by presenting subjects with stimuli in pairs and asking them to decide which of each pair looked 'older'. Measures like this provide information about relative-age judgments, but would have been too crude for our purposes. First, asking subjects to provide explicit age estimates for faces shows that they are capable of doing this to a high level of accuracy—something which would not be revealed by ranking methods. (This accuracy is particularly impressive given that in the case of an individual face, there is not necessarily a strong correspondence between chronological age and apparent age.)

Second, and perhaps more important, asking subjects to estimate ages is also more likely to reflect any disruptive effects of manipulations on age-estimation performance. Even with the Threshold versions in our experiments subjects were able to estimate age to some extent: note that, had we merely required subjects to rank order the stimuli in terms of apparent age, they would have been able to do this quite competently and we might well have concluded (erroneously) that the Threshold manipulation had no effect on their performance. Ranking measures show that subjects can estimate ages to some extent, but they provide insufficient insight into precisely how well people can do so.

The presence of pseudo-cardioidal strain in the Moved versions of the target faces led to subjects rating the faces as appearing younger than the Original versions. This is consistent with previous research demonstrating the effects of similar transformations on apparent age (eg Pittenger and Shaw 1975; Mark and Todd 1983; Bruce et al 1991), although the effects of the strain transformation in our experiments were weaker than the results of previous research might have led us to expect. One reason for this is that in experiment 2 we used a between-subjects design, which may have been less sensitive to the effects of strain than the within-subjects designs typically used by previous researchers to investigate this phenomenon. However, the effects of strain were relatively small in experiment 1, in which we did employ a within-subjects design.

Although there was a trend for Moved faces of all ages to appear younger than the corresponding Original faces, the results of experiment 1 implied that the effects of pseudo-cardioidal strain on age judgments were strongest for older subjects examining the younger target faces. In experiment 1, younger subjects' age estimates were significantly affected by pseudo-cardioidal strain for only the first age period (5 – 10-year-old target faces); in experiment 2, their estimates were largely unaffected by this manipulation even for this age period. Older subjects were more affected by pseudo-cardioidal strain at all ages in both experiments (although nonsignificantly so for some age categories in experiment 2). The fact that pseudo-cardioidal strain had greater effects on younger faces makes sense in developmental terms: cardioidal strain is a transformation which models craniofacial growth from infancy to adulthood; as Mark et al (1980) pointed out, since head growth does not continue indefinitely, cardioidal strain might be a useful cue to the age of children but it is unlikely to be used as a cue to the age of adults.

The relatively small effects of pseudo-cardioidal strain in our experiments may be due to an important methodological point. Studies such as those by Pittenger and his colleagues, in which cardioidal strain is the only cue present, can show that subjects are able to use this cue (although, as shown by our results for the Threshold condition, not particularly accurately); they do not show that this is the cue that people normally use when trying to assess the age of faces, because they do not give the subjects any option but to use cardioidal strain. Our Moved faces were relatively realistic and contained cues to age other than pseudo-cardioidal strain; the effects of these other cues may simply outweigh the effects of cardioidal strain in normal age perception.

Finally, it must be conceded that stronger effects of the cardioidal strain manipulation might have been obtained had we implemented this transformation in a mathematically rigorous way. Our attempts at producing pseudo-cardioidal strain by simply moving the features downwards in the face space may not have done justice to the subtle alterations in facial configuration induced by true cardioidal strain.

On the basis of our results, we suggest that age perception is more complex than previous research on cardioidal strain might lead one to believe. When one provides subjects with realistic pictures of faces and moves away from impoverished and artificial stimuli, such as line drawings and outline profiles, it becomes apparent that there are a variety of cues to age within a face and that subjects are adept at using them. If subjects are forced to use only one cue because only one is available (such as cardioidal strain or skin texture) it is still possible for them to make reasonably accurate age judgments but, in practice, cues to age are probably normally used in combination. This reinforces the conclusions drawn by Bruce et al (1989), who found that providing subjects with a richer database on which to base their age judgments (computer-generated 3-D head models) led to poorer and more idiosyncratic performance than that found by Mark and Todd (1985) who used simple line drawings of head profiles.

The differences in performance between our younger and older subjects suggest that it might be important to take the subjects' own age into account when researching this topic. In our experiments older subjects tended to be more heavily influenced by pseudo-cardioidal strain than did younger subjects; the reasons for this are not clear, but these results imply that in this kind of research one needs to take account of the age of the subject in relation to the age of the target. In experiment 1, older subjects tended to overestimate the age of younger faces and younger subjects tended to overestimate the age of older faces. Both groups were most accurate when estimating the age of a face within their own age range.

Our younger and older subjects obviously differ in many ways apart from just their age, but it is tempting to speculate that their performance differences stem from differential familiarity with faces of different ages. Differential familiarity with faces has been shown to affect processing in the domain of face recognition [eg the 'other-race effect' (Valentine 1991) and the effects of expertise (eg Diamond and Carey 1986; Rhodes and McLean 1990)], although the mechanisms responsible for these effects remain somewhat unclear. Whether differential familiarity is responsible for our results (and, if so, how), is an issue for further research. Of course there is also the more prosaic possibility that our observed differences are merely an artefact of the different ways in which our two age groups responded to the experimental demands. However, whatever the source of these differences, it is obviously desirable to be aware of the fact that results obtained from undergraduates on some aspects of face processing may have limited applicability to other groups.

Acknowledgements. We would like to thank Nicola Yuill and Richard Kemp for their comments on previous drafts of this paper, Tony Stubbins and Colin Crook for their technical assistance, and the children and adults who allowed us to use their faces as stimulus materials.

References

Bruce V, Burton M, Doyle T, Dench N, 1989 "Further experiments on the perception of growth in three dimensions" *Perception & Psychophysics* **46** 528 – 536

Bruce V, Doyle D, Dench T, Burton M, 1991 "Remembering facial configurations" *Cognition* **38** 109 – 144

Bruce V, Young A, 1986 "Understanding face recognition" *British Journal of Psychology* **77** 305 – 327

Diamond R, Carey S, 1986 "Why faces are not special: an effect of expertise" *Journal of Experimental Psychology: General* **115** 107 – 117

Enlow D H, 1982 *Handbook of Facial Growth* 2nd edition (Philadelphia, PA: W B Saunders Company)

Jones G, Smith P K, 1984 "The eyes have it: young children's discrimination of age in masked and unmasked facial photographs" *Journal of Experimental Child Psychology* **38** 328 – 337

Mark L S, Pittenger J B, Hines H, Carello C, Shaw R E, Todd J T, 1980 "Wrinkling and head shapes as coordinated sources of age-level information" *Perception & Psychophysics* **27** 117 – 124

Mark L S, Todd J T, 1983 "The perception of growth in three dimensions" *Perception & Psychophysics* **33** 193 – 196

Mark L S, Todd J T, 1985 "Describing perceptual information about human growth in terms of geometric invariants" *Perception & Psychophysics* **37** 245 – 256

Montepare J M, McArthur L Z, 1986 "The influence of facial characteristics on children's age perception" *Journal of Experimental Child Psychology* **42** 303 – 314

Pittenger J B, Shaw R E, 1975 "Ageing faces as visceral-elastic events: Implications for a theory of nonrigid shape perception" *Journal of Experimental Psychology: Human Perception and Performance* **1** 374 – 482

Pittenger J B, Shaw R E, Mark L S, 1979 "Perceptual information for the age-level of faces as a higher-order invariant of growth" *Journal of Experimental Psychology: Human Perception and Performance* **5** 478-493

Rhodes G, McLean I G, 1990 "Distinctiveness and expertise effects with homogeneous stimuli—towards a model of configural coding" *Perception* **19** 773-794

Shepherd J W, Davies G M, Ellis H D, 1981 "Studies of cue saliency", in *Perceiving and Remembering Faces* Eds G Davies, H Ellis, J Shepherd (London: Academic Press)

Smith A D, Winograd E, 1978 "Adult age differences in remembering faces" *Developmental Psychology* **14** 443-444

Valentine T, 1991 "A unified account of the effects of distinctiveness, inversion and race in face recognition" *Quarterly Journal of Experimental Psychology, Section A—Human Experimental Psychology* **43** 161-204